R E P O R T R E S U M E S

ED 011 169 FL 000 087

OBJECTIVE TESTING AND STANDARDIZED TESTS IN MODERN FOREIGN LANGUAGES.

BY- FAULKNER, JAMES C.

EDRS PRICE MF-$0.09 HC-$0.32 8P.

DESCRIPTORS- *BIBLIOGRAPHIES, *LANGUAGES, *LANGUAGE TESTS, *OBJECTIVE TESTS, *STANDARDIZED TESTS, ACHIEVEMENT TESTS, MULTIPLE CHOICE TESTS, TEST VALIDITY, TEST CONSTRUCTION,

A BRIEF HISTORY IS GIVEN OF THE DEVELOPMENT OF OBJECTIVE FOREIGN LANGUAGE TESTS FOR HIGH SCHOOLS AND COLLEGES, FROM 1916 TO THE PRESENT. THERE IS A DESCRIPTION OF THE EARLY EFFORTS TO CONSTRUCT OBJECTIVE TESTS AND THE IMPERFECTIONS OF SOME OF THESE EXAMINATIONS. LATER TESTS, PARTICULARLY THOSE IN CURRENT USE, ARE DESCRIBED IN DETAIL, WITH EMPHASIS ON THE VARIETY OF TECHNIQUES USED IN TESTING THE LANGUAGE SKILLS. STANDARDIZED TESTS CURRENTLY AVAILABLE FOR FRENCH, GERMAN, ITALIAN, RUSSIAN, AND SPANISH ARE LISTED, WITH ANNOTATIONS OF FORMAT, CONTENT, LEVEL, PRICE OF SPECIMEN SETS, AND TIME REQUIRED FOR ADMINISTRATION. (AM)

ED011169

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

OBJECTIVE TESTING AND STANDARDIZED TESTS IN MODERN FOREIGN LANGUAGES

By James C. Faulkner, Chairman
Division of Foreign Languages
State University College
New Paltz, New York

The impetus which brought forth the development of objective educational tests to measure achievement in subject matter can be traced back to the general intelligence scales of Binet in 1905, 1908 and 1911, to the Thorndike scale for measuring handwriting ability in 1910 and to the Army Alpha and Beta tests in 1917. The attempt, made by Professor Daniel Starch in 1916, to prepare and standardize an objective test to measure ability in Latin, French and German, is typical of the earliest efforts to apply these principles of objective testing to measure achievement in the study of foreign languages. In his book, Educational Measurements (New York: The Macmillan Co., 1917, Chapters XI, XII, and XIII, pages 171-187), he presents the tests he had constructed, together with norms for 1, 2, 3, and 4 years of study. Each test is composed of two parts. The Reading Test, which provides 30 sentences to be translated into English, can hardly be called objective. The Vocabulary Test utilizes the matching-term technique. It contains 100 words in the foreign language, arranged alphabetically, the English equivalents of which must be found among the 100 English words, also arranged alphabetically. Two such vocabulary lists are given because "an average obtained from two lists gives a very reliable score." The method of selection of items to be tested leaves much to be desired. The words were selected for the French lists by "taking the first word on every sixth or seventh page alternately in Spier's and Surenne's large French-English dictionary. List I was obtained by taking the first word on pages 6, 13, 19, 26, etc. List II was obtained by taking the first word on pages 7, 14, 20, 27, etc." Among the words beginning with B he chose balauder, batelage, bistortier, bouncher; under C he chose caique, carriole, cerf, chasser, choquer, cocuage, commissionaire, condylion, cophrophage, crasse, and cumul. The advantages of this method of selection, says Professor Starch are: "first, it gives a representative and uniform sampling of the entire --vocabulary. Second, the score obtained has a definite significance in the sense that it indicates the percentage of the entire --vocabulary that a person knows. If a pupil knows 25 words of each list, it means that he knows 25 per cent of the entire --vocabulary."

In view of the last statement, the teachers who administered the test for the purposes of standardization must have been quite elated over the achievement of their pupils, since the average score given in the norms for the first year high school students was 38. On first examining the French list, one is surprised at such a score, but after comparing the French list with the English list, one discovers at least 35 cognates which would have been correctly chosen by the average person who had never studied French, had the direction been given thus: "put in front of each French word the number of the English word which resembles it most closely as to spelling." Any person would have matched such words as analyse, anecdote, attentive, content, disputer, frontispiece, etc.

Among the other early tests were the Handschin Modern Language Tests in French and Spanish (1920), the Henmon French Tests (1921) and the Méras-Roth-Wood French Scales (1923). The defects of these early tests and the naivete of the claims for their reliability make us realize that, however imperfect the language tests in use today may be, they demonstrate that much progress has been made in the application of the principles of objective measurement.

FL 000 087

Interest in the development of reliable tests and dissatisfaction with the existing ones were best expressed in Dr. Ben D. Wood's extensive study of the Regent's Examinations in the report of the Modern Foreign Language Study (1925). The report was instrumental in preparing the way for a long series of standardized tests which provided measurements for vocabulary, silent reading, grammar, aural comprehension, knowledge of vowel and consonant values, of written comprehension, of civilization and culture; in fact, of all the aspects of the student's achievement and progress in learning a foreign language except the most important one: speaking.

Among the best known of these tests, mention must be made of the American Council Alpha and Beta Tests in French, German, and Spanish (1926), the Columbia Research Bureau Modern Language Tests in French and Spanish (1926), the Twigg French Vocabulary Test (1927), the Test of Spanish Vocabulary and the Silent Reading Test in Spanish by Contreras, Brown, and Kaulfers (1927), the American Council German Reading Scales (1927), the Silent Reading Test in French by Broom and Brown (1928), the Sammartino-Krause Standard French Tests (1928), and the Stanford Spanish Tests (1929). Though many of these tests were imperfect and had recognized limitations, extensive experiments by Wood, Ford, Cheydleur and others indicated that they had a greater reliability and certainly were more valid than the early ones.

The revision in 1934 of Professor Henmon's Achievement Tests in the Modern Foreign Languages (New York: The Macmillan Co., 1929) under the sponsorship of the Committee on Modern Languages of the American Council on Education marks a new era in foreign language testing. First there was the establishment of the Cooperative Test Service by the American Council and the creation by the Committee on Modern Languagesof new tests which have had widespread use until very recently. Important to note is the fairly general use of the various testing techniques which these new tests introduced especially in the comprehensive testing of reading.

These Cooperative Tests in French, German and Spanish used the multiple-response technique. Incomplete sentences in the foreign language had to be completed by one of five words or phrases, also in the foreign language. Each of the words or phrases were of the same level of difficulty as the words being tested and they all presented reasonable plausibility. For the more advanced items on those tests, a paragraph of text was given and statements based on it followed which were to be completed by one of five choices. The method certainly succeeded in measuring knowledge of vocabulary and idioms, and to a more limited degree, knowledge of advanced syntax.

The technique used by Cheydleur, Henmon and Walker in constructing the American Council French Reading Test was somewhat similar. A passage of continuous narrative was given to which questions were asked in the foreign language. The student had to select the best answer among the three given in English. The German American Council Reading Test contained ten German passages on which a total of 50 German questions were to be answered in English. The technique employed in the German test was undoubtedly easier to use than that of the French test but it lost something in objectivity since the student answered the questions in English and did not choose one among three answers given. It was necessary that each question be so phrased as to prevent, in the answers, any varying degrees of comprehension if each response were to be scored objectively, that is, right or wrong. The key also had to be prepared with extreme care.

The reading test developed at the University of Iowa by Foder and Cochran and described in the Modern Language Journal, January 1936, utilized a new technique, which was developed first to test Reading Comprehension in English. Called a Reading Comprehension Maturity Test, it was designed to test not

only the degree of comprehension of factual material but also the student's ability to evaluate the material before him, the depth of his understanding and the varying degrees of comprehension of complete thought-units. In this test a reading passage was provided with instructions to read through the entire passage before attempting to work the test items. The test items were arranged in groups of four statements based on a designated paragraph. One of these statements was false, at least in part, one indicated comprehension of some outstanding detail, one showed more complete understanding, and one summarized the essential meaning of the whole paragraph. The student was requested to indicate which statement was the expression of the most essential meaning of the paragraph and which one was the "worst or wrong" expression of the paragraph. The reliability co-efficient obtained for this test was usually high and it took better account of intelligence than most other reading tests. The results seemed to indicate that it did measure the individual differences in the ability to perceive the interrelationships of ideas.

Another technique used in testing reading comprehension by means of word-in-context identification was that developed by Professor Bond and others at the University of Chicago. The test was designed to measure the comprehension of words, idioms and grammar variables in all their contextual variations. A continuous reading passage was given with certain words and phrases underlined. These words and phrases were repeated below the passage, where each was followed by four suggested English translations. The student had to select the translation which was most appropriate in the context in which the words appeared.

With the recent developments in the approach to modern foreign language teaching and the emphasis given to the teaching of languages for communication, new tests have appeared in the past few years, the most interesting and valid ones, without doubt, being the MLA Cooperative Tests in French, German, Italian, Russian, and Spanish.

This is a new series of tests of competence designed for use in secondary schools and colleges. Forms available are: LA, LB for those having one to two full years of study in secondary school, or one or two semesters of study in colleges; MA, MB for those having three to four full years of study in secondary school, or three to four semesters of study in college. In the LISTENING test (25 minutes), the student listens to taped material and then answers multiple-choice questions based on: (1) single utterances, (2) short conversations between two speakers, (3) passages of connected discourse read by a single speaker, (4) telephone conversations for which the examinee selects the appropriate response, (5) dramatic scenes enacted by two to four speakers.

In the SPEAKING test (10 minutes), the student receives his instructions from a master tape at an individual recording station. He responds by speaking into a recorder. The test presents both verbal and visual stimuli. The student: (1) repeats what he hears with proper pronunciation and intonation, (2) reads aloud to demonstrate fluency, (3) responds to spoken questions involving a picture stimulus, (4) describes pictures presented singly or in sequence.

The READING test (35 minutes) is entirely a paper-and-pencil test, involving four-choice questions. (1) Sentence completion questions are used to measure the student's understanding of words and idiomatic expressions, (2) passages drawn from periodicals and appropriate literary sources are used to test word and phrase discrimination and ability to understand the main idea, find details and draw conclusions.

In the WRITING test (35 minutes), the student demonstrates his writing ability by actually writing both structured and free responses. (1) He fills in

blanks to show his understanding of correct usage of articles, prepositions, pronouns, verb forms, and simple idioms. In Russian, he also adds inflectional endings, (2) he rewrites sentences, making changes in tense, gender, number, person, word order, and sentence structure, (3) he writes or rewrites structured paragraphs and short dialogs based on verbal stimuli.

Scaled scores or percentile rankings are provided which permit the teacher to compare the work of his students with the national average for students at the same level of study. The results of these tests show the overlapping of achievement at the various levels and can be used for grouping the students according to ability. In fact, modern language teachers can feel assured that they can now obtain fairly accurate instruments for measuring the progress and final achievement of their students.

Some things remain still to be done, however. There should be a complete description of all published tests, including the levels at which a test is most valid, the technique employed, the reliability coefficients, norms, percentile rankings, scaled scores, the price and the publishers. The merits of the various tests should also be evaluated. Some standardized tests published many years ago are at present of little value compared to many of more recent date. Some have been withdrawn from publication, having been distributed at first in experimental form only. As a first step in that direction, we are giving below an incomplete listing of standardized tests in foreign languages presently available along with a short bibliography on evaluation of achievement in foreign languages.

Baltimore County French Test
Baltimore County Spanish Test

Level: One or two years of high school work
Both tests come in two parts (A and B) which can be used separately (40 minutes each).
Part I tests vocabulary, reading, grammar, culture, and aural comprehension. Part II tests aural comprehension exclusively through questions, identifications, and story. (Tape available) Answers are of multiple choice and true-false type. Specimen set: $ .50.
Bobbs-Merrill Co. Inc., 4300 West 62nd Street, Indianapolis 6, Ind.

Bureau of Educational Measurements (Kansas)

First Year French Test

Level: high school and college
Form A. Objective test of 90 items built to measure vocabulary, grammar, pronunciation, translation, and sight reading for thought content. (40 minutes) Specimen set: $ .95

Second Year French Test

Level: high school and college
Form A. Comprehensive test of 95 items in vocabulary, grammar, pronunciation, translation and sight reading for thought.
(40 minutes) Specimen set: $ .95

Kansas First-Year Spanish Test

Level: high school and college
Comprehensive test on vocabulary, grammar, pronunciation, translation, sight reading for thought, and some general cultural information.
Specimen set: $ .25

Kansas Second Year Spanish Test

Level: high school and college
Comprehensive test of 100 items on vocabulary, sentence and paragraph reading and comprehension, grammar idioms, translations, culture.
Specimen set: $ .25

Spanish and Latin American Life and Culture

Level: high school and college
This objective test consists of 100 items which covers geography, history, art, literature, and customs of Spain and Latin America (40 minutes) Specimen set: $ .25

Aiken-Held First Year German Test

Level: high school and college
Objective test high in validity and reliability Specimen set: $ .25

Bureau of Educational Measurements
Kansas State Teachers College
Emporia, Kansas

Bureau of Educational Research and Service

Iowa Placement Examinations

Forms available:
Foreign language Aptitude, Form A, B, M
French Training, Form A, B
Spanish Training, Form A, B
Speciment set: $ .20 each form, except Form M: $ .45

Bureau of Educational Research and Service
State University of Iowa
Iowa City, Iowa

Common Concepts Foreign Language Tests

Level: I (not limited to any particular grade)
Available in two forms in French, German, and Spanish. The basic approach is audio-lingual. The student hears stimulus sentences in a foreign language and indicates his understanding of what he has heard by selecting from a set of four small pictures the one which is correctly described. (40 minutes) Specimen set: $1.00

California Test Bureau
Del Monte Research Park
Monterey, California

Cooperative Foreign Language Tests

Level: Elementary: 2 years of high school or 1 year of college; Advanced: beyond 2 years of high school or 1 year of college. Objective tests similar for both levels. Part I tests reading through sentence-completion, and reading comprehension; Part II tests, vocabulary through synonyms; Part III tests grammar by means of English sentences partially translated in the foreign language;

student selects the missing item from five choices. (40 minutes)
Forms available:
French Form Q elementary, Form Q advanced.
French Form R elementary, Form R advanced.
Spanish Form O elementary, Form O advanced.
Spanish Form P elementary, Form P advanced.
Specimen set: $2.00

Cooperative French Listening Comprehension Test

Level: 2nd year high school through 5th semester college
Test developed to measure achievement in understanding spoken native French (30 minutes) Two forms (A and B) are available.
Specimen set: $2.00

Educational Testing Service
Princeton, New Jersey

Minnesota High School Achievement Examinations

Level: grades 7-12
The following forms are available (one hour)
French I and II
German I and II
Spanish I and II
Specimen set: $ .20 each

American Guidance Service, Inc.
Educational Test Bureau
720 Washington Avenue S. E.
Minneapolis 14, Minnesota

Ohio Scholarship Tests

Second Every Pupil Tests in German

Level: grades 9-12
Objective test (40 minutes) Specimen set: $ .05

District-State Scholarship Tests in French and Spanish

Level: grades 10-12
Objective tests intended for selected students. Two forms (First year and second year) for each language are available. Specimen set: $ .10

Division of Guidance and Testing, State Department of Education
751 Northwest Boulevard, Columbus 12, Ohio

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.